

# Algebraic Structures and Stochastic Differential Equations driven by Lévy processes

Charles Curry · Kurusch Ebrahimi–Fard ·  
Simon J.A. Malham · Anke Wiese

22nd December 2015

**Abstract** We define a new numerical integration scheme for stochastic differential equations driven by Lévy processes that has a leading order error coefficient less than that of the scheme of the same strong order of convergence obtained by truncating the stochastic Taylor series. This holds for all such equations where the driving processes possess moments of all orders and the coefficients are sufficiently smooth. The results are obtained using the quasi-shuffle algebra of multiple iterated integrals of independent Lévy processes. Our findings generalize recent results concerning equations driven by Wiener processes.

**Keywords** efficient integrators · Lévy processes · quasi-shuffle algebra

---

Charles Curry

Maxwell Institute for Mathematical Sciences and School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK.

*Present address:* of Charles Curry

Department of Mathematical Sciences, NTNU, 7491 Trondheim, Norway.

Kurusch Ebrahimi–Fard

Instituto de Ciencias Matemáticas, Consejo Superior de Investigaciones Científicas, C/Nicolás Cabrera, no. 13-15, 28049 Madrid, Spain.

Simon J.A. Malham

Maxwell Institute for Mathematical Sciences and School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK.

Anke Wiese

Maxwell Institute for Mathematical Sciences and School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK.

## 1 Introduction

Our main result is the derivation of a new strong numerical integration scheme for stochastic differential equations driven by independent Lévy processes with moments of all orders, where the coefficient functions are sufficiently smooth autonomous vector fields. Lévy processes are a class of stochastic processes, comprising processes continuous in probability and possessing stationary increments, independent of the past. They are examples of stochastic processes with a well understood structure that incorporate jump discontinuities.

Our setting is a complete, filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ , assumed to satisfy the usual hypotheses—see Protter [34, p.3]. Due to the Lévy decomposition of a Lévy process the equations we consider may be written in the form

$$y_t = y_0 + \sum_{i=0}^d \int_0^t V_i(y_s) dW_s^i + \sum_{i=d+1}^{\ell} \int_0^t V_i(y_{s-}) dJ_s^i.$$

See for example Applebaum [2]. Here the governing autonomous vector fields  $V_i: \mathbb{R}^N \rightarrow \mathbb{R}^N$  are supposed to be smooth and in general non-commuting, and the initial data  $y_0 \in L^2(\Omega, \mathcal{F}_0, P)$ . The  $W^i$  for  $i = 1, \dots, d$  are independent Wiener processes, and by convention we set  $W_t^0 = t$ . The  $J^i$  for  $i = d+1, \dots, \ell$  are purely discontinuous martingales expressible in the form

$$J_t^i = \int_0^t \int_{\mathbb{R}} v \bar{Q}^i(dv, ds),$$

where  $\bar{Q}^i(dv, ds) := Q^i(dv, ds) - \rho^i(dv)ds$ . The  $Q^i$  are Poisson measures on  $\mathbb{R} \times \mathbb{R}_+$  with intensity measures  $\rho^i(dv)ds$ , where the  $\rho^i$  are measures on  $\mathbb{R}$  with  $\rho^i(0) = 0$  and  $\int_{\mathbb{R}} (1 \wedge v^2) \rho^i(dv) < \infty$ . Intuitively,  $Q^i(B, (a, b])$  counts the number of jumps of the process  $J^i$  taking values in the set  $B$  in the time interval  $(a, b]$ , whilst  $\rho^i(B)$  measures the expected number of jumps per unit time the process  $J^i$  accrues taking values in the set  $B$ . We assume that the  $J^i$  possess moments of all orders. Lévy processes have many applications, notably in mathematical finance for the construction of models going beyond the Black–Scholes–Merton model in incorporating discontinuities in stock prices, see Cont & Tankov [8] and Barndorff–Nielsen, Mikosch & Resnick [3] and the references therein. There are also applications to physical sciences, see for instance Barndorff–Nielsen, Mikosch & Resnick [3]. Pathwise integrals of Lévy processes have recently been constructed in the framework of rough paths by Friz & Shekhar [15].

We are concerned with strong integration schemes for such Lévy-driven stochastic differential equations. These schemes are typically derived from the stochastic Taylor expansion. This is an expression for the flowmap  $\varphi_{s,t}: y_s \mapsto y_t$  as a sum of iterated integrals with an explicit remainder, which formally can be extended to give an infinite series. See Platen & Bruti-Liberati [32] for a derivation of the stochastic Taylor expansion for Lévy-driven stochastic differential equations, as well as Platen [30, 31] and Platen & Wagner [33].

For sufficiently smooth vector fields, integration schemes of arbitrary strong order of convergence may be constructed by truncating the series expansion to an appropriate number of terms. Platen & Bruti-Liberati [32] show which terms must be retained to obtain a strong integration scheme with a given strong order of convergence. We assume hereafter the governing vector fields are sufficiently smooth for the stochastic Taylor expansion to exist.

Combinatorial algebras are a natural setting for the study of relations among the iterated integrals appearing in stochastic Taylor expansions. In this paper we focus on the case when the stochastic Taylor expansion can be written in separated form. This means that each term in the stochastic Taylor expansion can be decomposed as the product of a multiple Itô integral, containing stochastic information only, and a composition of associated differential operators, containing geometric information only. In this instance the algebra of integrand-free multiple integrals of the driving processes is of interest alongside the algebra of the associated differential operators. Curry, Ebrahimi-Fard, Malham & Wiese [10] proved that the algebra of multiple integrals of a minimal family of semimartingales is isomorphic to the combinatorial quasi-shuffle algebra of words. A set of Lévy processes represents one example. Indeed Gaines [19] considered the case of multiple iterated integrals of Wiener processes and Li & Liu [22] considered multiple iterated integrals of Wiener processes and standard Poisson processes. The quasi-shuffle algebra is an extension of the shuffle algebra introduced abstractly, and comprehensively, by Hoffman [17]. The significance of the shuffle algebra was cemented in the work of Eilenberg & MacLane [14], Schützenberger [38] and Chen [6]. See Reutenauer [36] for more details.

One advantage of abstracting to the quasi-shuffle algebra is that we can immediately identify the minimum set of iterated integrals that need to be simulated to implement a given accurate strong scheme. This optimizes the total computation time which is dominated by the strong simulation of the iterated integrals. Radford [35] proved that the shuffle algebra is generated by Lyndon words. This was extended to the quasi-shuffle algebra by Hoffman [17], though it had already been established by Gaines [19] for the case of Wiener processes and Li & Liu [22] for the case of Wiener processes and standard Poisson processes, while Sussmann [40] had considered a Hall basis. Hence the set of iterated integrals we need to simulate at any given order are identified by Lyndon words. Having thus optimized the total computation time associated with a scheme of a given strong order, the focus turns to how to minimize the coefficient of the leading order error. This problem was considered by Clark [7] and Newton [28, 29] for drift-diffusion equations driven by a single Wiener process; also see Kloeden & Platen [20, Section 13.4]. They derived integration schemes that are asymptotically efficient in the sense that the coefficient of the leading order error is minimal among all schemes. For Stratonovich stochastic differential equations, Castell & Gaines [4, 5] constructed integration schemes using the Chen-Strichartz exponential Lie series expansion (see Strichartz [39]). In the strong order one case for one driving Wiener process,

and the strong order one-half case for two or more driving Wiener processes, their schemes are asymptotically efficient in the sense of Clark and Newton.

Malham & Wiese [25] considered the problem of designing efficient integrators for Stratonovich drift-diffusion equations driven by multiple independent Wiener processes. Here, efficient integrators of a given order have leading order error which is always less than that for the corresponding stochastic Taylor scheme, independent of the governing vector fields. They considered a class of strong integration schemes we shall hereafter call map-truncate-invert schemes. These are constructed as follows. Given an invertible map  $f: \text{Diff}(\mathbb{R}^N) \rightarrow \text{Diff}(\mathbb{R}^N)$ , we expand the series  $f(\varphi)$ , truncate according to a chosen grading and then apply  $f^{-1}$ . The case  $f = \log$  corresponds to the exponential Lie series integrator of Castell & Gaines. Malham & Wiese [25] then proved that an integration scheme based on the map  $f = \sinh\log$  is efficient. These results were obtained in the absence of a drift term and extended to incorporate drift in Ebrahimi-Fard, Lundervold, Malham, Munthe-Kaas & Wiese [11].

Herein, we study map-truncate-invert schemes for Itô equations driven by Lévy processes. We must consider the algebraic structure of iterated integrals with respect to Lévy processes, for which we rely on recent results of Curry, Ebrahimi-Fard, Malham & Wiese [10]. In this paper, we:

1. Show the stochastic Taylor series expansion for the flowmap can be written in separated form, i.e. as a series of terms, each of which is decomposable into a multiple Itô integral and a composition of associated differential operators;
2. Describe the class of map-truncate-invert schemes for Lévy-driven equations and give an algebraic framework for encoding and comparing such schemes. These results generalize those in Malham & Wiese [25] and Ebrahimi-Fard *et al.* [11];
3. Prove truncations according to the word length grading give more accurate schemes than approximations of the same order of convergence obtained by truncations according to the mean-square grading;
4. Introduce the antisymmetric sign reverse integrator, a new integration scheme for Lévy-driven equations, represented as half the difference of the identity and sign reverse endomorphisms on the vector space generated by words indexing multiple integrals. This scheme is efficient in the sense that its leading order mean-square error is less than that of the corresponding stochastic Taylor scheme, independent of the governing vector fields;
5. Show how to compute map-truncate-invert schemes in practice, in particular, how to deal with the inversion stage. We call these direct map-truncate-invert schemes. In essence their implementation involves computing the stochastic Taylor expansion to the given order, and then adding a polynomial of the already simulated multiple integrals to emulate the map-truncate-invert scheme.

The structure of this paper is as follows. In §2, we discuss the stochastic Taylor expansion for Lévy-driven equations given in Platen & Bruti-Liberati [32]. We

show when and how we can represent the stochastic Taylor expansion in separated form. In §3, we introduce map-truncate-invert schemes in the context of equations driven by Lévy processes. We show how to algebraically encode them using the quasi-shuffle convolution endomorphism algebra, thus extending the encoding for drift-diffusion equations in Ebrahimi–Fard et al. [11]. In §4 we undertake a local error analysis of integration schemes for Lévy-driven equations. A component of this analysis is the comparison of two gradings, the mean-square grading and the word length grading. Our main results are presented in §5. We prove that the antisymmetric sign reverse integrator, an extension of the sinhlog integrator of Malham & Wiese [25], is efficient. We also introduce our direct map-truncate-invert schemes. We demonstrate these with two numerical experiments in §6. Finally in §7 we derive global convergence results from local error estimates.

## 2 Separated stochastic Taylor expansions

A stochastic Taylor expansion is an expression for the flowmap as a sum of iterated integrals. We write down the stochastic Taylor expansion for Lévy-driven equations and show how and when it can be written in separated form. Recall the flowmap is defined as the map  $\varphi_{s,t}: y_s \mapsto y_t$ . It acts on sufficiently smooth functions  $f: \mathbb{R}^N \rightarrow \mathbb{R}^N$  as the pullback  $\varphi_{s,t}(f(y_s)) := f(y_t)$ . We set  $\varphi_t := \varphi_{0,t}$ . Itô's formula (see for example Applebaum [2, p. 203] or Protter [34, p. 71]) implies

$$f(y_t) = f(y_0) + \sum_{i=0}^d \int_0^t \tilde{V}_i \circ f(y_s) dW_s^i + \sum_{i=d+1}^{\ell} \int_0^t \int_{\mathbb{R}} (\tilde{V}_i \circ f)(y_{s-}, v) \bar{Q}^i(dv, ds),$$

where for  $i = 1, \dots, \ell$ , the  $\tilde{V}_i$  are operators defined as follows

$$\tilde{V}_i \circ f := \begin{cases} (V_i \cdot \nabla) f, & \text{if } i = 1, \dots, d, \\ f(\cdot + v V_i(\cdot)) - f(\cdot), & \text{if } i = d+1, \dots, \ell, \end{cases}$$

and  $\tilde{V}_0$  is the operator defined by

$$\begin{aligned} \tilde{V}_0 \circ f &:= (V_0 \cdot \nabla) f + \frac{1}{2} \sum_{i=1}^d \sum_{j,k=1}^N V_i^j V_i^k \partial_{x_j} \partial_{x_k} f \\ &\quad + \sum_{i=d+1}^{\ell} \int_{\mathbb{R}} [(\tilde{V}_i \circ f)(\cdot, v) - v(V_i \cdot \nabla) f] \rho^i(dv). \end{aligned}$$

Note for  $i = d+1, \dots, \ell$ , the  $\tilde{V}_i$  introduce an additional dependence on a real parameter  $v$ . The stochastic Taylor expansion is derived by expanding the integrands in the Lévy-driven equation using Itô's formula. This procedure is repeated iteratively, where the iterations are encoded as follows. Let  $A$  be the alphabet  $A := \{0, 1, \dots, \ell\}$ . For any such alphabet, we use  $A^*$  to denote the

free monoid over  $A$ —the set of words  $w = a_1 \dots a_m$  constructed from letters  $a_i \in A$ . We write  $\mathbb{1}$  for the empty word. For a given word  $w = a_1 \dots a_m$  define the operator  $\tilde{V}_w := \tilde{V}_{a_1} \circ \dots \circ \tilde{V}_{a_m}$ . Let  $s(w)$  be the number of letters of  $w$  from the subset  $\{d+1, \dots, \ell\} \subset A$ . For a given integrand  $g(t, v): \mathbb{R}_+ \times \mathbb{R}^{s(w)} \rightarrow \mathbb{R}^N$ , we define the iterated integrals  $I_w(t)[g]$  inductively as follows. We write  $I_{\mathbb{1}}(t)[g] := g(t)$ , and

$$I_w(t)[g] := \begin{cases} \int_0^t I_{a_1 \dots a_{m-1}}(s)[g] dW_s^{a_m}, & \text{if } a_m = 0, 1, \dots, d, \\ \int_0^t \int_{\mathbb{R}} I_{a_1 \dots a_{m-1}}(s_-)[g(\cdot, v)] \bar{Q}^{a_m}(dv, ds), & \text{if } a_m = d+1, \dots, \ell. \end{cases}$$

Iteratively applying the chain rule generates the following (see Platen & Bruti-Liberati [32]).

**Theorem 1 (Stochastic Taylor expansion)** *For a Lévy-driven equation the action of the flowmap  $\varphi_t$  on sufficiently smooth functions  $f: \mathbb{R}^N \rightarrow \mathbb{R}^N$  can be expanded as follows*

$$\varphi_t \circ f = \sum_{w \in A^*} I_w(t)[\tilde{V}_w \circ f].$$

*Remark 1 (Stochastic Taylor expansion convergence)* For integration schemes derived from the stochastic Taylor expansion to converge, it suffices that the terms  $\tilde{V}_w \circ f$  included in the expansion, and those at leading order in the remainder, satisfy global Lipschitz and linear growth conditions (see Platen & Bruti-Liberati [32]). Hereafter we assume these conditions are satisfied.

*Remark 2 (Platen and Bruti-Liberati form)* The stochastic Taylor expansion derived in Platen & Bruti-Liberati [32] is an equivalent though different representation. The equivalence of the expansions is seen from the identity

$$\int_0^t \int_{\mathbb{R}^{\ell-d}} (\tilde{V}_{-1} \circ f)(y_{s-}, v) \bar{Q}(dv, ds) = \sum_{i=d+1}^{\ell} \int_0^t \int_{\mathbb{R}} (\tilde{V}_i \circ f)(y_{s-}, v) \bar{Q}^i(dv, ds).$$

The expression on the left is the encoding employed by Platen and Bruti-Liberati of the jump terms in the stochastic differential equation while our encoding is that on the right. The equivalence is explained as follows. On the left above  $\bar{Q}$  is a compensated Poisson random measure on  $\mathbb{R}^{\ell} \times \mathbb{R}_+$ . The Lévy-driven equation may be written in this form, where  $V_{-1}(x, v) = \sum_{i=1}^{\ell-d} v_i V_{i+d}(x)$  and  $\bar{Q}$  is the compensation of the Poisson random measure  $Q$  defined such that  $Q(B, (a, b]) = \#\{\Delta J_s^1 \in B_1, \dots, \Delta J_s^{\ell} \in B_{\ell} : s \in (a, b]\}$  for a Borel set  $B = B_1 \times \dots \times B_{\ell} \subset \mathbb{R}^{\ell}$  bounded away from the origin, i.e.  $0 \notin \bar{B}$ . As independent Lévy processes almost surely never jump simultaneously (see Cont & Tankov [8, Theorem 5.3]), the measure  $Q$  is concentrated on sets of the form  $0 \times \dots \times 0 \times B_i \times 0 \times \dots \times 0$  with intensity measure  $\rho(0 \times \dots \times 0 \times B_i \times 0 \times \dots \times 0)dt = \rho^i(B_i)dt$ . The identity above thus follows. The stochastic

Taylor expansion given in Platen & Bruti-Liberati [32] is of the same form as the expansion we have given, but where the alphabet is instead  $\{-1, 0, \dots, d\}$ , and the operator associated to the letter  $-1$  is the multi-dimensional shift  $\tilde{V}_{-1}: f(y) \mapsto f(y + V_{-1}(y, v)) - f(y)$ .

Given the stochastic Taylor expansion for the flowmap in Theorem 1, we now show how we can write it in separated form. One component of the separated form are iterated integrals which are free in the sense of having no integrand. We define these abstractly for an arbitrary given alphabet for the moment, the reason for this will be apparent presently.

**Definition 1 (Free multiple iterated integrals)** Given a collection of stochastic processes  $\{Z_t^{a_i}\}_{a_i \in \mathbb{A}}$ , indexed by a given countable alphabet  $\mathbb{A}$ , free multiple iterated integrals take the form

$$I_w(t) := \int_{0 < \tau_1 < \dots < \tau_m < t} dZ_{\tau_1}^{a_1} \dots dZ_{\tau_m}^{a_m},$$

where  $w = a_1 \dots a_m$  are words in  $\mathbb{A}^*$ .

We need to augment  $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$  with a set of extended driving processes as follows. A key component in the characterization of the algebra generated by the vector space of free iterated integrals of the driving processes  $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$  are the compensated power brackets defined for each  $i \in \{d+1, \dots, \ell\}$  and for  $p \geq 2$  by

$$J_t^{i(p)} := \int_0^t \int_{\mathbb{R}} v^p \bar{Q}^i(dv, ds).$$

Equivalently we have  $J_t^{i(p)} = [J^i]^{(p)} - t \int_{\mathbb{R}} v^p \rho^i(dv)$ , where  $[J^i]^{(p)}$  is the  $p$ th order nested quadratic covariation bracket of  $J_t^i$ . Importantly, the compensated power brackets have the property that if  $J_t^{i(p)}$  is contained in the linear span of  $\{t, J_t^i, J_t^{i(2)}, \dots, J_t^{i(p-1)}\}$  for some  $p \geq 2$ , then  $J_t^{i(q)}$  is also in this linear span for all  $q \geq p$ . Hence inductively for  $p \geq 2$ , we augment our family  $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$  to include the compensated power brackets  $J_t^{i(p)}$  as long as they are not contained in the linear span of  $\{t, J_t^i, J_t^{i(2)}, \dots, J_t^{i(p-1)}\}$ .

By doing so, we obtain a possibly infinite family of stochastic processes. The iterated integrals of this extended family form the algebra generated by the iterated integrals of our driving processes. See Curry et al. [10] for further details. In summary we define our extended alphabet as follows.

**Definition 2 (Extended alphabet)** We define our alphabet  $\mathbb{A}$  to contain the letters  $0, 1, \dots, \ell$ , associated with  $\{W^0, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$ , and the additional letters  $i^{(p)}$  corresponding to any  $J_t^{i(p)}$  contained in the extended family as described above.

**Definition 3 (Separated stochastic Taylor expansion)** The flowmap for a Lévy-driven stochastic differential equation possesses a separated stochastic Taylor expansion if it can be written in the form

$$\varphi_t = \sum_{w \in \mathbb{A}^*} I_w \tilde{V}_w,$$

where  $\{I_w\}_{w \in \mathbb{A}^*}$  are the free iterated integrals associated to the extended driving processes and  $\tilde{V}_w = \tilde{V}_{a_1} \circ \cdots \circ \tilde{V}_{a_m}$  are operators indexed by words that compose associatively.

*Remark 3 (Separable cases)* For  $i \in \{1, \dots, d\}$  and  $j \in \{d+1, \dots, \ell\}$ , consider the term

$$I_{ji}[(\tilde{V}_{ji} \circ \text{id})(y_0)] = \int_0^t \left( \int_0^s \int_{\mathbb{R}} \left( V_i(y_0 + vV_j(y_0)) - V_i(y_0) \right) \bar{Q}^j(dv, d\tau) \right) dW_s^i,$$

in the general stochastic Taylor expansion. The double integral with respect to the random measure is a sum of compensated power bracket processes and thus separable, if the integrand  $V_i(y_0 + vV_j(y_0)) - V_i(y_0)$  is either a polynomial in the jump size  $v$  or it can be expressed as a power series in  $v$ . We see how this works in detail presently.

*Remark 4 (Separated expansion for jump-diffusion equations)* In the case of jump-diffusion equations, i.e. Lévy-driven equations for which all the discontinuous driving processes  $J^i$  are standard Poisson processes, the stochastic Taylor expansion is of the separated form. Indeed, as standard Poisson processes have jumps of size one only, the operators  $\tilde{V}_i$  with  $i = d+1, \dots, \ell$  do not introduce a dependence on an additional parameter. The integrands are thus constant across the range of integration, and hence the expansion is separated.

We can construct a separated stochastic Taylor expansion for the flowmap as follows. By Taylor expansion of the term  $f(y + vV_i(y))$  appearing in the shift  $\tilde{V}_i \circ f$ , we have

$$(\tilde{V}_i \circ f)(y, v) = \sum_{m \geq 1} v^m \tilde{V}_{i(m)} \circ f(y),$$

where we write  $V_i = (V_i^1, \dots, V_i^N)^T$  and  $f = (f^1, \dots, f^N)^T$ . We define the operators  $\tilde{V}_{i(m)}$  by

$$\tilde{V}_{i(m)} \circ f^j := \sum_{k \geq 1} \sum_{i_1 + \dots + i_k = m} \frac{1}{m!} V_i^{i_1} \cdots V_i^{i_k} \frac{\partial^m f^j}{\partial y^{i_1} \cdots \partial y^{i_k}},$$

where the  $i_j \in \mathbb{N}$ . The product in  $V_i^{i_1} \cdots V_i^{i_k}$  is multiplication in  $\mathbb{R}$ . We then have

$$I_i(t)[(\tilde{V}_i \circ f)(\cdot, v)] = \sum_{m \geq 1} \int_0^t (\tilde{V}_{i(m)} \circ f) dJ_s^{i(m)},$$



for  $i = d + 1, \dots, \ell$ . Inserting the above into the relation

$$I_w(t)[\tilde{V}_w \circ f] = I_{a_2 \dots a_m}(t) \left[ I_{a_1}(\cdot) [\tilde{V}_{a_1} \circ (\tilde{V}_{a_2 \dots a_m} \circ f)] \right]$$

and iterating gives the separated expansion, where the operators  $\tilde{V}_a$  are those of the stochastic Taylor expansion for  $a \in \{0, 1, \dots, d\}$ , and for  $a = i^{(m)}$  are given by the  $\tilde{V}_{i^{(m)}}$  defined above. We have thus just established the following result.

**Theorem 2 (Separated stochastic Taylor expansion: existence)** *For a Lévy-driven equation suppose the terms  $\tilde{V}_w \circ f$  in the stochastic Taylor expansion for the flowmap are analytic on  $\mathbb{R}^N$ . Then it can be written in separated form, i.e. as a separated stochastic Taylor expansion.*

*Remark 5 (Linear vector fields and linear diffeomorphisms)* The separated expansion has a simple form when the governing vector fields are linear with constant coefficients and we consider the action of the flowmap on homogeneous linear diffeomorphisms, say  $f(y) = Fy$ , where  $F = [f_{ij}]$  is an  $N \times N$  matrix. The identity map is a special case of such a linear diffeomorphism which generates the solution  $y_t$  directly. In this case, the operators in the expansion are given by matrix multiplication. We write  $V_i(y) = A^i y$ , where  $A^i = [a_{jk}^i]$  are constant  $N \times N$  matrices and consider the action of the flowmap on linear functions. First, for  $i = 1, \dots, d$  by direct computation we find  $\tilde{V}_i \circ f(y) = FA^i y$ . Moreover, for  $i = d + 1, \dots, \ell$  the higher order differential operators  $\tilde{V}_{i^{(m)}}$  with  $m \geq 2$  vanish due to the linearity, and we obtain  $(\tilde{V}_i \circ f)(y, v) = v(V_i \cdot \nabla)f(y) = vFA^i y$ . The above relation shows in addition that the term in  $\tilde{V}_0 \circ f$  involving an integral over the jump sizes vanishes. As the functions we act on are linear, the second order terms of  $\tilde{V}_0$  also vanish. We therefore have  $\tilde{V}_0 \circ f(y) = f(V_0(y)) = FA^0 y$ . The operators  $\tilde{V}_i$  act by matrix multiplication. It follows that the operators  $\tilde{V}_w$  act on linear functions by matrix multiplication in the reverse order,  $\tilde{V}_{a_1 \dots a_k} \circ f(y) = FA^{a_k} \dots A^{a_1} y$ .

Hereafter we assume the existence of a separated Taylor expansion for the flowmap, and all iterated integrals are free iterated integrals.

### 3 Convolution algebras and map-truncate-invert schemes

We now introduce a class of numerical integration schemes we call map-truncate-invert schemes and show how they can be encoded algebraically. The algebraic structures arise naturally from the products of iterated integrals and the composition of operators appearing in the separated stochastic Taylor expansion. Consider the class of numerical integration schemes obtained from the stochastic Taylor expansion by simulating truncations of the expansion on each subinterval of a uniform discretization of a given time domain  $[0, T]$ .

**Definition 4 (Grading function and truncations)** A grading function  $g: \mathbb{A}^* \rightarrow \mathbb{N}$  assigns a positive integer to each non-empty word  $w \in \mathbb{A}^*$  and zero to the empty word. A truncation is specified by a grading function and truncation value  $n \in \mathbb{N}$ . We write  $\pi_{g=n}$ ,  $\pi_{g \leq n}$  and  $\pi_{g \geq n}$  for the projections of  $\mathbb{A}^*$  onto the following subsets:

1.  $\pi_{g=n}(\mathbb{A}^*) := \{w \in \mathbb{A}^* : g(w) = n\}$ ;
2.  $\pi_{g \leq n}(\mathbb{A}^*) := \{w \in \mathbb{A}^* : g(w) \leq n\}$ ; and
3.  $\pi_{g \geq n}(\mathbb{A}^*) := \{w \in \mathbb{A}^* : g(w) \geq n\}$ .

A numerical integration scheme based on the stochastic Taylor expansion is thus given by successive applications of the approximate flow

$$\sum_{w \in \pi_{g \leq n}(\mathbb{A}^*)} I_w(t) \tilde{V}_w$$

across the computational subintervals. Now more generally, consider a larger class of numerical schemes that are constructed as follows. For a given invertible map  $f: \text{Diff}(\mathbb{R}^N) \rightarrow \text{Diff}(\mathbb{R}^N)$ , we construct a series expansion for  $f(\varphi_t)$  using the stochastic Taylor expansion for  $\varphi_t$ . We truncate the series and simulate the retained iterated Itô integrals across each computational subinterval. An integration scheme is obtained by computing the inverse map  $f^{-1}$  of the simulated truncations at each step; see Malham & Wiese [25] and Ebrahimi-Fard *et al.* [11]. We now develop an algebraic framework for studying such map-truncate-invert schemes. The starting point is the quasi-shuffle algebra of Hoffman [17]. This gives an explicit description of the algebra of iterated integrals of the extended driving processes. Let  $\mathbb{R}\mathbb{A}$  denote the  $\mathbb{R}$ -linear span of  $\mathbb{A}$ .

**Definition 5 (Quasi-shuffle product)** For a given alphabet  $\mathbb{A}$ , suppose  $[\cdot, \cdot]: \mathbb{R}\mathbb{A} \otimes \mathbb{R}\mathbb{A} \rightarrow \mathbb{R}\mathbb{A}$  is a commutative, associative product on  $\mathbb{R}\mathbb{A}$ . The quasi-shuffle product on  $\mathbb{R}\langle \mathbb{A} \rangle$ , which is commutative, is generated inductively as follows: if  $\mathbb{1}$  is the empty word then  $u * \mathbb{1} = \mathbb{1} * u = u$  and

$$ua * vb = (u * vb)a + (ua * v)b + (u * v)[a, b],$$

for all words  $u, v \in \mathbb{A}^*$  and letters  $a, b \in \mathbb{A}$ . Here  $ua$  denotes the concatenation of  $u$  and  $a$ .

*Remark 6 (Word-to-integral isomorphism)* The word-to-integral map  $\mu: w \mapsto I_w$  is an algebra isomorphism. Here the domain is the vector space  $\mathbb{R}\langle \mathbb{A} \rangle$  equipped with the quasi-shuffle product. Iterated integrals indexed by polynomials are defined by linearity, i.e.  $I_{k_u u + k_v v} = k_u I_u + k_v I_v$ , for any two constants  $k_u, k_v \in \mathbb{R}$  and words  $u, v \in \mathbb{A}^*$ . This was proved in Curry *et al.* [10], it had already been established by Gaines [19] for drift-diffusions and by Li & Liu [22] for jump-diffusions.

*Remark 7 (Shuffle product)* The quasi-shuffle product is a deformation of the shuffle product on the vector space  $\mathbb{R}\langle \mathbb{A} \rangle$  of polynomials in the non-commuting

variables  $\mathbb{A}$ . The deformation is induced by an additional product  $[\cdot, \cdot]$  defined on  $\mathbb{R}\mathbb{A}$ . If the product  $[\cdot, \cdot]$  is trivial, the quasi-shuffle product is just the shuffle product.

*Remark 8 (Quadratic covariation bracket)* Here we associate the additional product  $[\cdot, \cdot]: \mathbb{R}\mathbb{A} \otimes \mathbb{R}\mathbb{A} \rightarrow \mathbb{R}\mathbb{A}$  underlying the quasi-shuffle with the pullback under the word-to-integral map  $\mu$  of the quadratic covariation bracket of semimartingales. Explicitly,  $[0, a]$  is zero for all  $a$ , and  $[i^{(p)}, j^{(q)}] = \delta_{ij}(\lambda(i, p, q) \cdot 0 + (1_{\{d+1, \dots, \ell\}}(i)) \cdot i^{(p+q)})$ , where  $\lambda(i, p, q) = 1$  if  $i \in \{1, \dots, d\}$ , and  $\lambda(i, p, q) = \int_{\mathbb{R}} v^{p+q} \rho^i(dv)$  if  $i \in \{d+1, \dots, \ell\}$ . The 0 refers to the letter  $0 \in \mathbb{A}$ , and  $1_{\{d+1, \dots, \ell\}}$  is the indicator function of the set  $\{d+1, \dots, \ell\}$ .

*Remark 9 (Word-to-operator homomorphism)* The word-to-operator map denoted by  $\kappa: w \mapsto \tilde{V}_w$  is an algebra homomorphism. Here the domain is  $\mathbb{R}\langle \mathbb{A} \rangle$  equipped with the concatenation product. This follows as  $\mathbb{R}\langle \mathbb{A} \rangle$  equipped with the concatenation product is the free associative  $\mathbb{R}$ -algebra over  $\mathbb{A}$ , see Reutenauer [36], and each  $\tilde{V}_w$  is given by the associative composition of operators  $\tilde{V}_{a_i}$ .

We write  $\mathbb{R}\langle \mathbb{A} \rangle_*$  for the quasi-shuffle algebra, otherwise the product on  $\mathbb{R}\langle \mathbb{A} \rangle$  is concatenation. The preceding observations combine to give an encoding of integration schemes in the algebra  $\mathbb{R}\langle \mathbb{A} \rangle_* \bar{\otimes} \mathbb{R}\langle \mathbb{A} \rangle$ , where  $\bar{\otimes}$  is the completed tensor product of Reutenauer [36, p. 18, 29].

**Proposition 1 (Algebraic encoding of the flowmap)** *For a given Lévy-driven equation, the map  $\mu \otimes \kappa$  is a homomorphism from  $\mathbb{R}\langle \mathbb{A} \rangle_* \bar{\otimes} \mathbb{R}\langle \mathbb{A} \rangle$  to the tensor product of the algebra of iterated integrals of the extended driving processes and the composition algebra generated by the set of operators  $\tilde{V}_w$ . Moreover, the flowmap of the equation is the image under  $\mu \otimes \kappa$  of the element*

$$\sum_{w \in \mathbb{A}^*} w \otimes w.$$

Truncations of this abstract series representation in  $\mathbb{R}\langle \mathbb{A} \rangle_* \bar{\otimes} \mathbb{R}\langle \mathbb{A} \rangle$  generate approximations of the flowmap and hence classes of stochastic Taylor numerical integration schemes. Using this context, we now construct an abstract representation for  $f(\varphi_t)$  for any given map  $f: \text{Diff}(\mathbb{R}^N) \rightarrow \text{Diff}(\mathbb{R}^N)$  which is expressible as a power series. The key idea is that  $f(\varphi_t)$  may be rewritten as the image under  $\mu \otimes \kappa$  of a series  $\sum F(w) \otimes w$ , where  $F \in \text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$ , the space of  $\mathbb{R}$ -linear maps from the quasi-shuffle algebra  $\mathbb{R}\langle \mathbb{A} \rangle_*$  to itself; see Malham & Wiese [25] who established this in the shuffle product context. The explicit form of  $F \in \text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$  is obtained using the convolution algebra associated to the quasi-shuffle product.

**Definition 6 (Quasi-shuffle convolution product)** For a given quasi-shuffle product  $*$  on  $\mathbb{R}\langle \mathbb{A} \rangle_*$ , the convolution product  $\star$  of  $F$  and  $G$  in  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$  is given by

$$F \star G := * \circ (F \otimes G) \circ \Delta,$$

where  $\Delta$  is the deconcatenation coproduct that sends a word  $w$  to the sum of all its two-partitions  $u \otimes v$ , i.e. explicitly for any word  $w \in \mathbb{A}^*$  we have  $\Delta(w) = \sum_{uv=w} u \otimes v$  and

$$(F \star G)(w) = \sum_{uv=w} F(u) * G(v).$$

The quasi-shuffle algebra  $\mathbb{R}\langle \mathbb{A} \rangle_*$  together with the coproduct  $\Delta$  forms a bialgebra, see Hoffman [17]. In particular, when equipped with the convolution product, the space  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$  becomes a unital algebra, where the unit  $\nu$  is given by the composition of the unit of the quasi-shuffle algebra and the counit of the deconcatenation coalgebra, see Abe [1]. Explicitly,  $\nu$  is the linear map that sends non-empty words to zero and the empty word to itself. We define the embedding  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*) \rightarrow \mathbb{R}\langle \mathbb{A} \rangle_* \overline{\otimes} \mathbb{R}\langle \mathbb{A} \rangle$  by

$$F \mapsto \sum_{w \in \mathbb{A}^*} F(w) \otimes w.$$

This is an algebra homomorphism for the quasi-shuffle convolution product; see Reutenauer [36], Curry [9] and Ebrahimi-Fard *et al.* [13]. Given a power series  $f(x) = \sum_{k \geq 0} c_k x^k$  with  $c_k \in \mathbb{R}$ , we define the convolution power series  $F^*(X) := \sum_{k \geq 0} c_k X^{*k}$ , where the  $X^{*k}$  are the  $k$ th convolution powers of  $X \in \text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$ . The following representation of  $f(\varphi_t)$  is immediate.

**Proposition 2 (Convolution power series)** *If  $f$  has a power series  $f(x) = \sum_{k \geq 0} c_k x^k$ , we have*

$$f\left(\sum_{w \in \mathbb{A}^*} w \otimes w\right) = \sum_{w \in \mathbb{A}^*} F^*(\text{id})(w) \otimes w.$$

*In other words we can represent  $f(\varphi_t)$  by  $F^*(\text{id})$ . Furthermore, when the power series  $f$  has an inverse  $f^{-1}(x) = \sum_{k \geq 0} b_k x^k$  with  $b_k \in \mathbb{R}$ , the compositional inverse of the convolution power series for  $F$  is given by the associated convolution power series  $F^{-1}$ .*

Combining the algebraic encoding of the flowmap in Proposition 1 with Proposition 2, we construct the map-truncate-invert scheme associated to a given power series  $f$  as follows.

**Procedure 1 (Map-truncate-invert scheme)** *Let  $f: \text{Diff}(\mathbb{R}^N) \rightarrow \text{Diff}(\mathbb{R}^N)$  be an invertible map admitting an expansion as a power series. For a given truncation function  $\pi_{g \leq n}$ , the associated map-truncate-invert scheme across a fixed computational interval  $[0, t]$  is obtained as follows.*

1. Construct the series  $F^*(\text{id})$ ;
2. Simulate the truncation  $\pi_{g \leq n} \circ F^*(\text{id})$  given by

$$\hat{o}_t := \sum_{w \in \pi_{g \leq n}(\mathbb{A}^*)} I_{F^*(\text{id})(w)}(t) \tilde{V}_w(\text{id}).$$

3. Compute the approximation  $f^{-1}(\hat{\sigma}_t) \circ y_0$ .

*Remark 10* In  $\hat{\sigma}_t$  in (ii) above: The identity map in  $F^*(\text{id})(w)$  is the identity endomorphism in  $\text{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$ , while the identity map in  $\tilde{V}_w(\text{id})$  is the identity diffeomorphism on  $\mathbb{R}^N$ .

*Remark 11 (Map-truncate-invert endomorphism)* Map-truncate-invert schemes are realizations of the integration scheme associated to the endomorphism  $F^{-1} \circ \pi_{g \leq n} \circ F^*(\text{id}) \in \text{End}(\mathbb{R}\langle\mathbb{A}\rangle)$ .

An endomorphism which will prove useful in what follows is the augmented ideal projector.

**Definition 7 (Augmented ideal projector)** The augmented ideal projector denoted by  $\mathfrak{J}$  is given by  $\mathfrak{J} := \text{id} - \nu$ . In other words, it acts as the identity on non-empty words, but sends the empty word to zero.

*Example 1 (Exponential Lie series)* Recall from the introduction that the motivating example of a map-truncate-invert scheme was the exponential Lie series integrator, corresponding to  $f = \log$ . For any endomorphism  $X$  that maps the empty word to itself,  $\log^*(X)$  is a power series in  $X - \nu$ . In particular, we have  $\log^*(\text{id}) = \mathfrak{J} - \mathfrak{J}^{*2}/2 + \dots + (-1)^{k+1}\mathfrak{J}^{*k}/k + \dots$ .

*Remark 12* For any word  $w$ , say of length  $k$ , we can naturally truncate a series in convolutional powers of  $\mathfrak{J}$  with real coefficients  $c_1, c_2, c_3$  and so forth, to  $c_1\mathfrak{J} + c_2\mathfrak{J}^{*2} + \dots + c_k\mathfrak{J}^{*k}$ . This is because the action of  $\mathfrak{J}^{*(k+1)}$  and subsequent terms in the series is, by convention, zero as a word of length  $k$  cannot be partitioned into more than  $k$  non-empty parts.

The computation of  $f^{-1}(\hat{\sigma}_t)$  from  $\hat{\sigma}_t$  is in general non-trivial. For the exponential Lie series integrator, Castell & Gaines [4,5] proposed computing  $\exp(\hat{\sigma}_t)(y_0)$  by numerical approximation of the ordinary differential equation  $u' = \hat{\sigma}_t(u)$ , subject to the initial condition  $u(0) = y_0$ . An approximate solution may then be recovered from  $u(1)$ . For the sinhlog integrator in the shuffle product context, Malham & Wiese [25] showed that for linear constant coefficient equations, since  $\hat{\sigma}$  is a square matrix, the approximate flow  $\exp \sinh^{-1}(\hat{\sigma}_t) = \hat{\sigma}_t + (\text{id} + \hat{\sigma}_t^2)^{1/2}$  may be computed using a matrix square root. If the vector fields are nonlinear Malham & Wiese [25] suggested expanding the square root to sufficiently high degree terms. In §5 we introduce direct map-truncate-invert schemes which evaluate  $F^{-1} \circ \pi_{g \leq n} \circ F^*(\text{id})$  directly and circumvent the step involving the computation of  $f^{-1}$  in general.

## 4 Endomorphism inner product, gradings and error analysis

We establish convergence and accuracy of integration schemes at an algebraic level. We first define an inner product on the space of endomorphisms corresponding to the  $L^2$  inner product of the associated approximate flows. We

then define the mean-square grading and introduce stochastic Taylor schemes of a specific local and then global order. We subsequently consider word length grading. We show stochastic Taylor integration schemes of a given strong order obtained by truncating according to word length are always more accurate than those obtained by truncating according to the mean-square grade. We measure accuracy by the leading order term in the  $L^2$ -norm of the remainder after truncation. To start, to define the endomorphism inner product, we require an algebraic encoding of the expectation of the iterated integrals.

**Definition 8 (Expectation map)** For any word  $w \in \mathbb{A}^*$ , the expectation map  $E: \mathbb{R}\langle \mathbb{A} \rangle \rightarrow \mathbb{R}[t]$  is defined by  $E: w \mapsto t^{|w|}/|w|!$  if  $w \in \{0\}^*$  and is zero for all other words. Here  $|w|$  denotes the length of the word  $w$ ,  $\{0\}^* \subset \mathbb{A}^*$  is the free monoid over the letter 0 and  $\mathbb{R}[t]$  is the polynomial ring over a single indeterminate  $t$  commuting with  $\mathbb{R}$ .

*Remark 13 (Expectation of iterated Itô integrals)* The expectation map corresponds to the expectation of iterated Itô integrals as follows. First, integrals indexed by words not ending in the letter 0 are martingales and hence have zero expectation. Second, consider integrals indexed by a word with at least one non-zero letter. Fubini's Theorem implies such integrals also have zero expectation; see Protter [34].

We now define an inner product on  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$ , following Ebrahimi-Fard *et al.* [11]. Suppose we apply two separate functions to the flowmap which are characterised by the endomorphisms  $F$  and  $G$  in  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$ . The stochastic processes generated by these functions of the flowmap, for given initial data  $y_0$ , are  $\sum_{w \in \mathbb{A}^*} I_{F(w)} \tilde{V}_w(y_0)$  and  $\sum_{w \in \mathbb{A}^*} I_{G(w)} \tilde{V}_w(y_0)$ . Our goal is to define an inner product  $\langle F, G \rangle$  on  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$  which matches the  $L^2$ -inner product of these two vector-valued stochastic processes. However we would like all of our results to be independent of the initial data  $y_0$  and the governing vector fields appearing in the  $\tilde{V}_w$  terms. We achieve this by replacing the vectors  $\tilde{V}_w(y_0)$  with a set of indeterminate vectors indexed by words,  $\{\mathbf{V}_w\}_{w \in \mathbb{A}^*}$ . We write  $(u, v)$  for the inner product of  $\mathbf{V}_u$  and  $\mathbf{V}_v$ , i.e.  $(u, v) := \mathbf{V}_u^T \mathbf{V}_v$ . Let  $\mathbf{V}$  denote the infinite square matrix indexed by the words  $u, v \in \mathbb{A}^*$  with entries  $(u, v)$ .

**Definition 9 (Inner product)** We define the inner product of endomorphisms  $F$  and  $G$  in  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$  with respect to  $\mathbf{V}$  to be

$$\langle F, G \rangle := \sum_{u, v \in \mathbb{A}^*} E(F(u) * G(v)) (u, v).$$

All results we subsequently establish will hold independent of  $\mathbf{V}$ .

*Remark 14 (Positive definiteness)* As the operators  $\tilde{V}_i$  typically include second (or higher) order differential operators that send the identity map to zero, distinct endomorphisms may be associated with the same stochastic process. For linear constant coefficient equations for instance, the operators  $\tilde{V}_{i(m)}$  are zero operators for all  $m > 1$ , and the stochastic process associated to any

endomorphism is trivial if its image lies in the two-sided ideal in  $\mathbb{R}\langle\mathbb{A}\rangle_*$  generated by the letters  $i^{(m)}$ ,  $m > 1$ . To obtain positive definiteness of the inner product, we pass to the quotient of  $\text{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$  under the equivalence relation for which endomorphisms yielding the same stochastic process are equivalent. Positive definiteness on the quotient space follows by similar arguments to those in Ebrahimi–Fard et al. [11].

The norm, in this quotient space, of an endomorphism  $F \in \text{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$  is  $\|F\| := \langle F, F \rangle^{1/2}$ .

The first part of our error analysis uses mean-square grading, see Platen & Bruti-Liberati [32].

**Definition 10 (Mean-square grading)** For any word  $w \in \mathbb{A}^*$ , the map  $g^{\text{ms}}: w \mapsto 2\zeta(w) + \xi(w)$  is the mean-square grading, where  $\zeta(w)$  and  $\xi(w)$  are the number of zero and non-zero letters in  $w$ , respectively.

**Definition 11 (Reduced words)** For a given word  $w$  the reduced word  $\text{red}(w)$  is defined as the word obtained by deleting any zero letters, and replacing any letters of the form  $i^{(m)}$  with  $i$ .

*Example 2* If  $w = 010024^{(3)}30$ , we have  $\text{red}(w) = 1243$ .

The following result will be useful in our subsequent error analysis, see Kloeden & Platen [20].

**Lemma 1 (Expectation of products)** Let  $\mathbb{R}\langle\mathbb{A}\rangle_*$  be the quasi-shuffle algebra based on the set of independent Lévy processes  $\{t, W^1, \dots, W^d, J^{d+1}, \dots, J^\ell\}$  extended by covariation. For any words  $u, v \in \mathbb{A}^*$ , if  $\text{red}(u) = \text{red}(v)$  then there is a non-zero constant  $C = C(u, v)$  such that

$$E(u * v) = C(u, v) t^{(g^{\text{ms}}(u) + g^{\text{ms}}(v))/2},$$

otherwise the expectation of the product  $u * v$  is zero.

*Proof* Let  $\langle p, w \rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$  denote the coefficient of a given word  $w$  in the polynomial  $p$ . We then have

$$E(u * v) = \sum_{w \in \{0\}^*} \langle u * v, w \rangle_{\mathbb{R}\langle\mathbb{A}\rangle} \frac{1}{|w|!} t^{|w|}.$$

Consider the generating relation of the quasi-shuffle product in Definition 5, which states that  $ua * vb = (u * vb)a + (ua * v)b + (u * v)[a, b]$ . We see that if a summand in the polynomial  $ua * vb$  is to be a multiple of  $w \in \{0\}^*$ , we require either  $a = 0$ ,  $b = 0$  or  $\langle [a, b], 0 \rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$  non-zero. As the quadratic covariation of  $t$  with any Lévy process vanishes, all zero letters contribute to the expectation only through the first two terms in the quasi-shuffle generating relation above. In contrast, non-zero letters  $a$  and  $b$  contribute only through the third term. We see that each zero letter appears exactly once in any  $w \in \{0\}^*$  for which  $\langle u * v, w \rangle_{\mathbb{R}\langle\mathbb{A}\rangle}$  is non-zero, and any pair of non-zero letters  $a, b$  contribute one

letter together. In particular, we have  $E(u * v) = C(u, v)t^{(g^{ms}(u) + g^{ms}(v))/2}$ , where  $C(u, v)$  may equal zero.

By the independence of the driving processes, we have  $[i, j]$ ,  $[i^{(p)}, j]$  and  $[i^{(p)}, j^{(q)}]$  are all zero for all  $i, j, p, q$ ,  $i \neq j$ . Moreover, we have  $[J^{i^{(p)}}, J^{i^{(q)}}]_t = J_t^{i^{(p+q)}} + t \int_{\mathbb{R}} v^{p+q} \rho^i(dv)$ . In particular, we have that  $\langle [i^{(p)}, i^{(q)}], 0 \rangle_{\mathbb{R}\langle \mathbb{A} \rangle}$  is non-zero for all  $i > d$  and all  $p, q \geq 1$  for which  $i^{(p)}$  and  $i^{(q)}$  exist. Moreover, as  $[W^i, W^i]_t = t$ , we have  $[i, i] = 0$  for  $i = 1 \dots, d$ .

The expression  $u * v$  is a linear combination of words, each of which arises from a choice of one of the three terms at each stage in the inductive quasi-shuffle generating relation. Indeed the  $k$ th letter in a given word thus obtained is the letter  $a$ ,  $b$  or  $[a, b]$  chosen at the  $k$ th application of the inductive definition. For this word to consist of only zeros, we must choose the first or the second term as long as either  $a$  or  $b$  is zero. When  $a$  and  $b$  are both non-zero, we must choose the third term. From the preceding paragraph, this will be a sum featuring a multiple of the zero letter provided both letters are either equal to  $i$ , where  $i = 1, \dots, d$ , or are  $i^{(p)}$  and  $i^{(q)}$  respectively for some  $i > d$  and  $p, q \geq 1$ . Continuing this procedure, we see that we will obtain a word comprising only zeros if and only if the reduced words  $\text{red}(u)$  and  $\text{red}(v)$  are equal.  $\square$

We wish to compare the errors of different integration schemes across a given time step at the algebraic level, using the inner product defined above. As the flowmap corresponds to the identity in  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$ , we define the remainder endomorphism as follows.

**Definition 12 (Remainder endomorphism)** For any endomorphism  $H \in \text{End}(\mathbb{R}\langle \mathbb{A} \rangle_*)$  encoding an approximation of the flowmap, we define the associated remainder endomorphism  $R$  to be

$$R := \text{id} - H.$$

*Example 3* A simple example is  $H = \pi_{g \leq n}$  corresponding to a stochastic Taylor expansion truncated according to a given grading ‘g’. In this case  $R = \text{id} - \pi_{g \leq n} = \pi_{g \geq n+1}$ . More generally we might have  $H = F^{-1} \circ \pi_{g \leq n} \circ F(\text{id})$  so  $R = \text{id} - F^{-1} \circ \pi_{g \leq n} \circ F(\text{id})$ .

For the rest of this section we focus on numerical schemes constructed by truncating the stochastic Taylor expansion, first, according to mean-square grading and, second, according to word length, grading. We note the flowmap  $\varphi_t \in L^2$ , see Lemma 5. Let  $\hat{\varphi}_t^{\text{ms}}$  denote the truncated stochastic Taylor expansion, truncated according to mean-square grading. We denote the corresponding remainder by  $R_t^{\text{ms}} := \varphi_t - \hat{\varphi}_t^{\text{ms}}$  and note  $R_t^{\text{ms}} \in L^2$ , see Platen and Bruti-Liberati [32]. Locally the numerical scheme is of mean-square order  $n$  if

$$R_t^{\text{ms}} = \sum_{w \in \pi_{g \geq n+1}(\mathbb{A}^*)} I_w(t) \tilde{V}_w.$$



In particular this implies, for sufficiently small  $t$ , that  $\|R_t^{\text{ms}}(y_0)\|_{L^2}^2 = (1 + |y_0|^2) \cdot \mathcal{O}(t^{n+1})$  for any initial data  $y_0 \in \mathbb{R}^N$ . We naturally apply the numerical scheme  $\hat{\varphi}_t^{\text{ms}}$  successively over a suitably fine discretization of the global time interval of integration to obtain a suitably accurate numerical approximation. This means we need to determine the global order of convergence for such a scheme. Milstein's Theorem [26, 27] provides a mechanism for inferring global convergence estimates from local ones in the case of drift-diffusion equations. This can be extended to Lévy driven equations and is provided in Theorem 5 in §7. We call it the Generalized Milstein Theorem. To apply this theorem we require local expectation estimates for  $R_t^{\text{ms}}(y_0)$ . We find that

$$E(R_t^{\text{ms}}(y_0)) = \sum_{k \geq \lfloor n/2 \rfloor + 1} \frac{t^k}{k!} \tilde{V}_{0^k}(y_0).$$

Here we used that  $E(I_w(t))$  is only non-zero for words  $w \in \{0\}^*$ . The notation  $0^k$  denotes such a word  $w$  of length  $k$ , and  $\text{g}^{\text{ms}}(0^k) = 2k$ . Using the linear growth estimates we observe for some constant  $K > 0$  we have

$$|E(R_t^{\text{ms}}(y_0))| \leq K(1 + |y_0|^2)^{1/2} \left( \sum_{k \geq \lfloor n/2 \rfloor + 1} \frac{t^k}{k!} \right).$$

For any finite  $t$  the sum is convergent. In particular for small  $t$  the upper bound is  $\mathcal{O}(t^{\lfloor n/2 \rfloor + 1})$ . Recall from above that  $\|R_t^{\text{ms}}(y_0)\|_{L^2}^2 = (1 + |y_0|^2) \cdot \mathcal{O}(t^{n+1})$ . We now refer to the Generalized Milstein Theorem 5 in §7. Matching parameters we see that  $p_1 = \lfloor n/2 \rfloor + 1$  and  $p_2 = (n + 1)/2$ . The theorem states that the approximation  $\hat{\varphi}_t^{\text{ms}}$  with remainder  $R_t^{\text{ms}}$  above will converge globally at rate  $p_2 - 1/2$  if  $p_1 \geq p_2 + 1/2$ . While this is true for when  $n$  is even, it does *not* hold when  $n$  is odd. This is simply due to the fact that pure deterministic terms in the stochastic Taylor series remainder have a whole integer less root mean-square global order of convergence compared to their local order of convergence. To rectify this we can simply modify our scheme  $\hat{\varphi}_t^{\text{ms}}$  to

$$\hat{\varphi}_t^{\text{ms}} = \sum_{\text{g}^{\text{ms}}(w) \leq n} I_w(t) \tilde{V}_w + I_{0^{n^*}}(t) \tilde{V}_{0^{n^*}},$$

where  $n^* := \lfloor (n + 1)/2 \rfloor$  if  $n$  is odd and zero if it is even. This means that the leading order deterministic term in  $R_t^{\text{ms}}$  is of the same order as previously when  $n$  is even, but is of order  $\lfloor (n + 1)/2 \rfloor + 1 = \lfloor (n + 3)/2 \rfloor$  when  $n$  is odd. By inspection we observe that  $p_1 \geq p_2 + 1/2$  is now satisfied. The modified scheme  $\hat{\varphi}_t^{\text{ms}}$  above has mean-square global order of convergence  $n$  and the terms included exactly match those specified in Platen and Bruti-Liberati [32, p. 290].

We now consider word length grading which we employ for our main result in the next section.

**Definition 13 (Word length grading)** For a given word  $w \in \mathbb{A}^*$ , the word length grading is denoted  $|w|$  or  $\text{g}^{\text{wl}}$ , and defined to be the number of letters in  $w$ .

*Remark 15 (Computational effort)* The bulk of the computation effort, when implementing accurate strong numerical schemes derived from truncations of series representations of the flowmap, is associated with the simulation of the iterated integrals  $I_w$ . See Lord, Malham and Wiese [23] and Malham and Wiese [24] for more details in the drift-diffusion case. In particular the iterated integrals involving the most distinct non-deterministic letters require the most effort. Hence the additional computational cost required to simulate all the iterated integrals  $\{I_w : |w| \leq n\}$  as opposed to the subset  $\{I(w) : g^{\text{ms}}(w) \leq n\}$ , is minimal.

One benefit of truncating to word length as opposed to mean-square grading is the following.

**Theorem 3 (Mean-square versus word length graded truncations)**

*Consider two approximate flowmaps arising from the truncation of the stochastic Taylor expansion, one using mean-square grading and the other using word length grading, both truncated at the same given grade  $n$ . The  $L^2$  norm of the leading order local flow remainder associated with the word length approximation is always less than that of the mean-square approximation.*

*Proof* First, recall the definitions for the mean-square and word length gradings. We observe for any word  $w \in \mathbb{A}^*$  we have  $g^{\text{ms}}(w) = 2\zeta(w) + \xi(w)$  and  $g^{\text{wl}}(w) = \zeta(w) + \xi(w)$ . Hence we have  $g^{\text{wl}} \leq g^{\text{ms}}$ . This implies that for any alphabet  $\mathbb{A}$  constructed from at least one deterministic and one non-deterministic letter, we have  $\pi_{g^{\text{ms}} \leq n}(\mathbb{A}^*) \subset \pi_{g^{\text{wl}} \leq n}(\mathbb{A}^*)$ . Thus correspondingly for their complements  $\pi_{g^{\text{wl}} \geq n+1}(\mathbb{A}^*) \subset \pi_{g^{\text{ms}} \geq n+1}(\mathbb{A}^*)$ . Second, let  $R^{\text{ms}}$  and  $R^{\text{wl}}$  denote the remainder endomorphisms associated with truncating the stochastic Taylor expansion by mean-square and word length gradings, respectively. Hence for the inner products  $\langle R^{\text{ms}}, R^{\text{ms}} \rangle$  and  $\langle R^{\text{wl}}, R^{\text{wl}} \rangle$  we sum over words in  $\pi_{g^{\text{ms}} \geq n+1}(\mathbb{A}^*)$  and  $\pi_{g^{\text{wl}} \geq n+1}(\mathbb{A}^*)$ , respectively. The difference remainder endomorphism  $\hat{R} := R^{\text{ms}} - R^{\text{wl}}$  is at leading order non-zero on words in  $\pi_{g^{\text{ms}} = n+1}(\mathbb{A}^*) \cap \pi_{g^{\text{wl}} \leq n}(\mathbb{A}^*)$ . For any word  $w$  in this set we have  $2\zeta(w) + \xi(w) = n + 1$  and  $\zeta(w) + \xi(w) \leq n$ , which imply  $\zeta(w) > 0$  and thus  $\xi(w) < n + 1$ . Now consider the inner product  $\langle R^{\text{wl}}, \hat{R} \rangle$ . Any word  $u$  on which  $R^{\text{wl}}$  is non-trivial at leading order has length  $\zeta(u) + \xi(u) \geq n + 1$ . Any word  $v$  on which  $\hat{R}$  is non-trivial at leading order, we have just shown that  $\xi(v) < n + 1$ . Hence we must have  $\xi(u) \neq \xi(v)$  and so  $\text{red}(u) \neq \text{red}(v)$ . We deduce that  $\langle R^{\text{wl}}, \hat{R} \rangle$  is zero. The result then follows from the relation  $\|R^{\text{ms}}\|^2 = \|R^{\text{wl}}\|^2 + 2\langle \hat{R}, R^{\text{wl}} \rangle + \|\hat{R}\|^2$ . This result holds whether we include the  $0^{n^*}$  in  $\hat{\varphi}_t^{\text{ms}}$  or not.  $\square$

## 5 Antisymmetric sign reverse integrator

We now present the error analysis of map-truncate-invert schemes, concluding with the description of the antisymmetric sign reverse integrator and the main

theorem concerning its efficiency. We also present the direct map-truncate-invert schemes alluded to at the end of §3. We begin by introducing the pre-remainder endomorphism associated to any map-truncate-invert scheme, that is the remainder terms associated to the truncation of the series  $F^*(\text{id})$ . We relate the pre-remainder and remainder of such schemes and use this relation to compute the inverse step of a direct map-truncate-invert scheme. We follow with the introduction of the antisymmetric sign reverse integrator and its explicit characterization. We then prove our main theorem on the efficiency of the antisymmetric sign reverse integrator.

**Definition 14 (Pre-remainder endomorphism)** Let  $F: \text{End}(\mathbb{R}\langle\mathbb{A}\rangle_*) \rightarrow \text{End}(\mathbb{R}\langle\mathbb{A}\rangle_*)$  be an invertible map and suppose  $\pi_{g \leq n}$  is a projection corresponding to a truncation. The pre-remainder endomorphism  $Q$  associated to  $\pi_{g \leq n} \circ F(\text{id})$  is defined by

$$Q := F(\text{id}) - \pi_{g \leq n} \circ F(\text{id}).$$

This definition allows for more general maps  $F$  than power series in the convolution algebra, which we require presently. The relationship between the pre-remainder and remainder is critical to the error analysis of map-truncate-invert schemes.

Hereafter we assume a grading  $g$  which is preserved by the quasi-shuffle product in question. By this we mean that for all words  $u, v \in \mathbb{A}^*$ , the polynomial  $u * v$  is homogeneous of degree  $g(u) + g(v)$ , i.e. it is a sum of words of grade  $g(u) + g(v)$ ; see Ebrahimi-Fard et al. [11]. In this case we observe that if  $\pi$  represents one of the projectors  $\pi_{g \leq n}$ ,  $\pi_{g=n}$  or  $\pi_{g \geq n}$  according to such a grading  $g$ , then  $\pi \circ \mathfrak{J}^{*k} = \mathfrak{J}^{*k} \circ \pi$  for any  $k \in \mathbb{N}$ .

*Remark 16 (Gradings preserved by quasi-shuffles)* The power bracket grading introduced in Curry et al. [10, Section 4], defined on uncompensated brackets, is grading preserving for any quasi-shuffle. This assigns the value 2 to the letter 0, the value 1 to letters  $1, \dots, \ell$ , and the sum of the gradings for the quadratic covariation bracket between any two words. Mean-square grading is preserved for drift-diffusions, however it is not preserved in general for any quasi-shuffle containing discontinuous terms. Importantly, word length grading  $g^{\text{wl}}$  is preserved when the quasi-shuffle product is in fact the shuffle product. This latter case will underlie our eventual application below.

**Lemma 2 (Remainder and pre-remainder relation)** Suppose  $f(1+x) = \sum_{k \geq 1} c_k x^k$  is a power series with inverse  $f^{-1}(x) = 1 + \sum_{k \geq 1} b_k x^k$ , where necessarily  $b_1 \equiv 1/c_1$ . Let  $\pi_{g \leq n}$  be the projection according to the grading preserved by the quasi-shuffle product. Let  $R$  and  $Q$  be the remainder and pre-remainder, respectively, associated to the map-truncate-invert endomorphism  $F^{-1} \circ \pi_{g \leq n} \circ F^*(\text{id})$ . Then we have

$$R = \frac{1}{c_1} Q + \text{h.o.t..}$$

*Remark 17 (Higher order terms)* The notation ‘h.o.t.’ in Lemma 2 is used to denote higher order terms in the following sense. Suppose our goal is to apply the endomorphism  $G^*(\text{id}) = C_1\mathfrak{J} + C_2\mathfrak{J}^{*2} + C_3\mathfrak{J}^{*3} + \dots$  to all words  $w$  with  $g(w) \leq n$ . If we write  $G^*(\text{id}) = C_1\mathfrak{J} + \dots + C_n\mathfrak{J}^{*n} + \mathcal{O}(H)$ , this implies the endomorphism  $H$  annihilates all words  $w$  with  $g(w) \leq n$ . In this sense the term(s) in  $\mathcal{O}(H)$  represent higher order terms. For example, suppose for a particular word  $w$  we have  $g(w) = n$ . The term  $\mathfrak{J}^{*(n+1)}$  splits  $w$  into a sum of all its possible non-empty  $(n+1)$ -partitions quasi-shuffled together of the form  $w_1 * w_2 * \dots * w_{n+1}$ . Since the grading is preserved by the quasi-shuffle product, we have  $g(w_1) + \dots + g(w_{n+1}) = n$ . However since the partitions  $w_i$  for all  $i = 1, \dots, n+1$  are all non-empty so that  $g(w_i) \geq 1$ , we have a contradiction. By convention we suppose  $\mathfrak{J}^{*(n+1)}$  annihilates such a word  $w$  and so  $\mathfrak{J}^{*(n+1)}$  represents higher order term in the sense we have outlined.

*Proof* We set  $P := \pi_{g \leq n} \circ F^*(\text{id})$ . Then we see the pre-remainder  $Q = \pi_{g \geq n+1} \circ F^*(\text{id})$  and thus also  $P + Q = F^*(\text{id})$ . Third we observe that

$$\begin{aligned} R &= F^{-1}(P + Q) - F^{-1}(P) \\ &= \sum_{k \geq 1} b_k ((P + Q)^{*k} - P^{*k}) \\ &= b_1 Q + b_2 (P \star Q + Q \star P) + \text{h.o.t.} \end{aligned}$$

Fourth, we have  $P \star Q = (\pi_{g \leq n} \circ F^*(\text{id})) \star (\pi_{g \geq n+1} \circ F^*(\text{id}))$ . Since  $F^*(\text{id}) = c_1 \mathfrak{J} + \text{h.o.t.}$  we get  $P \star Q = (c_1 \pi_{g \leq n} \circ \mathfrak{J}) \star (\pi_{g \geq n+1} \circ (c_1 \mathfrak{J} + \dots + c_{n+1} \mathfrak{J}^{*(n+1)})) + \text{h.o.t.}$ . Hence we deduce that  $P \star Q$  annihilates any words of grade less than  $n+2$ . Thus the term  $b_2 P \star Q$ , as well as similarly  $b_2 Q \star P$ , represent higher order terms. Using that  $b_1 \equiv 1/c_1$  gives the result.  $\square$

A similar calculation was used to establish efficiency of the sinhlog integrator in Ebrahimi-Fard et al. [11]. We use it to compute the inverse step of a map-truncate-invert scheme as follows.

**Corollary 1 (Direct map-truncate-inverse schemes)** *Suppose  $f(1+x) = \sum_{k \geq 1} c_k x^k$  is an invertible power series and let  $\pi_{g \leq n}$  be a truncation according to the grading preserved by the quasi-shuffle product. Then  $F^{-1} \circ \pi_{g \leq n} \circ F^*(\text{id})$ , up to higher order terms, is given by*

$$\pi_{g \leq n} + \pi_{g = n+1} \circ \left( \mathfrak{J} - \frac{1}{c_1} F^*(\text{id}) \right).$$

*Proof* Starting with the definition of the remainder endomorphism  $R$ , and then subsequently using our results above, we observe

$$\begin{aligned}
F^{-1} \circ \pi_{g \leq n} \circ F^*(\text{id}) &= \text{id} - R \\
&= \pi_{g \leq n} + \pi_{g \geq n+1} - \frac{1}{c_1} Q + \text{h.o.t.} \\
&= \pi_{g \leq n} + \pi_{g \geq n+1} \circ \left( \text{id} - \frac{1}{c_1} F^*(\text{id}) \right) + \text{h.o.t.} \\
&= \pi_{g \leq n} + \pi_{g=n+1} \circ \left( \mathfrak{J} - \frac{1}{c_1} F^*(\text{id}) \right) + \text{h.o.t.}
\end{aligned}$$

In the last step we used that  $\mathfrak{J}$  is the identity on non-empty words.  $\square$

*Remark 18 (Direct map-truncate-invert in practice)* Note that using the power series representation for  $F^*(\text{id})$  the direct map-truncate-inverse scheme in Corollary 1 is given by

$$\begin{aligned}
\pi_{g \leq n} - \frac{1}{c_1} \pi_{g=n+1} \circ (c_2 \mathfrak{J}^{*2} + c_3 \mathfrak{J}^{*3} + \dots) \\
= \pi_{g \leq n} - \frac{1}{c_1} (c_2 \mathfrak{J}^{*2} + \dots + c_{n+1} \mathfrak{J}^{*(n+1)}) \circ \pi_{g=n+1}.
\end{aligned}$$

We observe in this formula the action of direct map-truncate-invert schemes. We simulate the corresponding truncated stochastic Taylor scheme represented by  $\pi_{g \leq n}$  and add the corresponding additional terms shown. Note the additional terms  $\mathfrak{J}^{*2}$ ,  $\mathfrak{J}^{*3}$  and so forth only involve products over lower order multiple Itô integrals that have already been simulated in  $\pi_{g \leq n}$ . For instance, for a word  $w$  with  $g(w) = n+1$ , the term  $\mathfrak{J}^{*2}(w)$  is the sum of all products of Itô integrals of the form  $I_u I_v$  with non-empty words  $u$  and  $v$  such that  $uv = w$  and using grade preservation,  $g(u) + g(v) = n+1$ . In particular  $g(u) \leq n$  and  $g(v) \leq n$ , so  $I_u$  and  $I_v$  are already included in  $\pi_{g \leq n}$ . The question arises as to whether a given direct map-truncate-inverse scheme converges. For the integrator we construct presently, this will be an automatic consequence of the convergence of the corresponding stochastic Taylor integrator according to word length grading.

We turn our attention to establishing our main result concerning an efficient integrator for equations driven by Lévy processes. For convenience we introduce the following bracketing notation for the bracket  $[\cdot, \cdot]$  in the quasi-shuffle product. For letters  $a$  we set  $[a] := a$ , while for words  $w = a_1 \dots a_k$  we set  $[w] := [a_1, [a_2, \dots, [a_{k-1}, a_k] \dots]]$ . The commutativity of the bracket means the order of the letters  $a_i$  is irrelevant and then its associativity means that all  $k$ -fold brackets are equivalent to the canonical form of left-to-right bracketing shown.

**Definition 15 (Reversal, sign reversal and quasi-shuffle antipode endomorphisms)** We define three endomorphisms on  $\mathbb{R}\langle \mathbb{A} \rangle$  as follows. If  $a_i \in \mathbb{A}$  for  $i = 1, \dots, n$  are letters, then we define the:

1. Reversal map:  $|S|: a_1 \dots a_n \mapsto a_n \dots a_1$ ;
2. Sign reversal map:  $S: a_1 \dots a_n \mapsto (-1)^n a_n \dots a_1$ ; and
3. Quasi-shuffle antipode:  $\hat{S}: a_1 \dots a_n \mapsto (-1)^n \sum [u_1] [u_2] \dots [u_k]$ ,

where for the quasi-shuffle antipode the sum is over all possible factorizations of  $a_n \dots a_1 = u_1 u_2 \dots u_k$  into non-empty subwords  $u_i$  for all  $k = 1, \dots, n$ .

*Remark 19 (Antipode and quasi-shuffle Hopf algebra)* The quasi-shuffle antipode is the quasi-shuffle convolution reciprocal map of the identity, i.e.  $\text{id} \star \hat{S} = \hat{S} \star \text{id} = \nu$ . Indeed, with this in hand, the quasi-shuffle algebra  $\mathbb{R}\langle \mathbb{A} \rangle_*$ , with deconcatenation  $\Delta$  as coproduct, is also a Hopf algebra. These results were proved by Hoffman [17].

*Remark 20* For the shuffle algebra where the bracket  $[\cdot, \cdot]$  is trivial, we have  $\hat{S} = S$ .

Malham & Wiese [25] and Ebrahimi-Fard et al. [11] considered drift-diffusion equations interpreted in the Stratonovich sense—which corresponds to the shuffle product case. The scheme of main interest therein was the map-truncate-invert integration scheme generated from the map  $f = \sinh\log$ . The shuffle convolution power series associated to the series  $f(1+x) = \sinh\log(1+x) = x + \frac{1}{2} \sum_{k \geq 2} (-1)^{k-1} x^k$  is expressible in the form  $F^{\sqcup}(X) = \frac{1}{2}(X - X^{\sqcup(-1)})$ . Here  $\sqcup$  represents the convolution product corresponding to the shuffle product. In the shuffle case,  $\text{id} \sqcup S = S \sqcup \text{id} = \nu$  and we therefore have the identity  $\sinh\log^{\sqcup}(\text{id}) = \frac{1}{2}(\text{id} - S)$ , and similarly  $\cosh\log^{\sqcup}(\text{id}) = \frac{1}{2}(\text{id} + S)$ . The efficiency results of Malham & Wiese [25] and Ebrahimi-Fard et al. [11] for the  $\sinh\log$  integrator rely on the identity  $\langle \pi_{g=n} \circ \sinh\log^{\sqcup}(\text{id}), \pi_{g=n} \circ \cosh\log^{\sqcup}(\text{id}) \rangle = 0$ . Here  $\langle \cdot, \cdot \rangle$  is an inner product on  $\text{End}(\mathbb{R}\langle \mathbb{A} \rangle_{\sqcup})$  similar to  $\langle \cdot, \cdot \rangle$  but where the underlying expectation map differs due to the use of Stratonovich integrals. The above result utilizes the shuffle algebra structure of multiple Stratonovich integrals with respect to continuous semimartingale integrators. To obtain an appropriate extension of the  $\sinh\log$  integrator to Lévy-driven equations, we broaden our definition of map-truncate-invert schemes. In particular, we introduce the antisymmetric sign reverse integrator as follows.

**Definition 16 (Antisymmetric sign reverse integrator)** The antisymmetric sign reverse integrator is the direct map-truncate-invert scheme associated with  $\frac{1}{2}(\text{id} - S)$ , with truncation according to word length.

*Remark 21* We cannot in general give an expression for  $\frac{1}{2}(\text{id} - S)$  as a convolution power series in  $\mathfrak{J}$  in any quasi-shuffle convolution algebra with non-trivial bracket. To see this, compare the action of  $S$  and a map  $\alpha\mathfrak{J} + \beta\mathfrak{J}^{\star 2}$  on a general word of length two. We see that  $S(ab) = ba$  and the identity  $(\alpha\mathfrak{J} + \beta\mathfrak{J}^{\star 2})(ab) = (\alpha + \beta)ab + \beta ba + \beta[a, b]$ . Comparing these two expressions, we see that  $-\alpha = \beta = 1$ , and that  $[a, b] = 0$  for all  $a, b$ . This implies that  $[\cdot, \cdot]$  must be trivial.

The following representation generates a practical implementation of the antisymmetric sign reverse integrator. In particular, it includes the inverse step.

**Lemma 3 (Direct representation of the antisymmetric sign reverse integrator)** *The antisymmetric sign reverse integrator is given by*

$$\sum_{|w| \leq n} I_w(t) \tilde{V}_w + \sum_{|w|=n+1} I_{\coshlog^{\sqcup}(w)}(t) \tilde{V}_w,$$

or equivalently

$$\sum_{|w| \leq n} I_w(t) \tilde{V}_w + \sum_{|w|=n+1} \left( I_{\coshlog^*(\text{id})(w)}(t) + \frac{1}{2} I_{(S-\hat{S})(w)}(t) \right) \tilde{V}_w.$$

*Proof* The map  $\frac{1}{2}(\text{id} - S)$  is identified with  $\sinhlog^{\sqcup}(\text{id})$ . The shuffle product preserves the word length grading so  $\pi_{g=n+1} \circ S = S \circ \pi_{g=n+1}$  and thus  $\pi_{g=n+1} \circ \sinhlog^{\sqcup}(\text{id}) = \sinhlog^{\sqcup}(\text{id}) \circ \pi_{g=n+1}$ . Hence applying Corollary 1 with  $F = \sinhlog^{\sqcup}$  and using that  $c_1 = 1$  in this case and the identity  $\sinhlog^{\sqcup}(\text{id}) + \coshlog^{\sqcup}(\text{id}) = \text{id}$ , we have

$$\begin{aligned} & (\sinhlog^{\sqcup})^{-1} \circ \pi_{g \leq n} \circ \sinhlog^{\sqcup}(\text{id}) \\ &= \pi_{g \leq n} + (\text{id} - \sinhlog^{\sqcup}(\text{id})) \circ \pi_{g=n+1} + \text{h.o.t.} \\ &= \pi_{g \leq n} + (\coshlog^{\sqcup}(\text{id})) \circ \pi_{g=n+1} + \text{h.o.t.} \\ &= \pi_{g \leq n} + (\coshlog^*(\text{id})) \circ \pi_{g=n+1} \\ &\quad + (\coshlog^{\sqcup}(\text{id}) - \coshlog^*(\text{id})) \circ \pi_{g=n+1} + \text{h.o.t.} \\ &= \pi_{g \leq n} + (\coshlog^*(\text{id})) \circ \pi_{g=n+1} + \frac{1}{2}(S - \hat{S}) \circ \pi_{g=n+1} + \text{h.o.t.}, \end{aligned}$$

using that  $\coshlog^{\sqcup}(\text{id}) = \frac{1}{2}(\text{id} + S)$  and  $\coshlog^*(\text{id}) = \frac{1}{2}(\text{id} + \hat{S})$ . Ignoring higher order terms, using the convolution algebra embedding and applying  $\mu \otimes \kappa$  then gives the desired result.  $\square$

*Remark 22 (Antisymmetric sign reverse integrator in practice)* The terms given by  $(\coshlog^*(\text{id})) \circ \pi_{g=n+1}$  and  $\frac{1}{2}(S - \hat{S}) \circ \pi_{g=n+1}$  correspond to polynomials of words of lower word length  $n$  or less. For the former term this is straightforward following analogous arguments to those in Remark 18—noting that the projection operator  $\pi_{g=n+1}$  is applied before  $\coshlog^*(\text{id})$ . For the latter term,  $(S - \hat{S}) \circ (a_1 \dots a_{n+1})$  consists of a sum of terms, each of which contains at least one bracket. For example one term would be  $(-1)^{n+1} [a_{n+1}, a_n] a_{n-1} \dots a_1$  which is a word of length  $n$ , and so forth. Hence the additional terms  $(\coshlog^*(\text{id})) \circ \pi_{g=n+1}$  and  $\frac{1}{2}(S - \hat{S}) \circ \pi_{g=n+1}$  required for implementing the direct antisymmetric sign reverse integrator consist of lower order Itô integrals we have already simulated to construct the stochastic Taylor integrator  $\pi_{g \leq n}$ .

*Example 4* Consider the order one antisymmetric sign reverse approximation. Since  $\coshlog^*(\text{id}) = \nu + \frac{1}{2}\mathfrak{J}^{*2} - \frac{1}{2}\mathfrak{J}^{*3} + \dots$  and  $(S - \hat{S})(a_1 a_2 a_3) = \frac{1}{2}([a_3 a_2] a_1 +$

$a_3[a_2a_1] + [a_3a_2a_1]$ ) this antisymmetric sign reverse approximation has the form

$$\sum_{|w| \leq 2} I_w \tilde{V}_w + \sum_{w=a_1a_2a_3} \frac{1}{2} (I_{a_1}I_{a_2a_3} + I_{a_1a_2}I_{a_3} - I_{a_1}I_{a_2}I_{a_3} + I_{[a_3a_2]a_1} + I_{a_3[a_2a_1]} + I_{[a_3a_2a_1]}) \tilde{V}_w.$$

**Lemma 4** *Let  $\pi_{g=n}$  be the projection according to the word length grading. Then for any  $\mathbf{V}$  we have:*

1.  $\langle X, Y \rangle = \langle |S| \circ X, |S| \circ Y \rangle;$
2.  $\|\pi_{g=n} \circ S\|^2 = \|\pi_{g=n} \circ \text{id}\|^2.$

*Proof* As  $E(w) = 0$  unless  $w \in \{0\}^*$ , we have  $E(w) = E(|S|(w))$ . By Hoffman & Ihara [18, Prop. 4.2], the reversal map is an automorphism for any quasi-shuffle algebra. Hence we have  $E(|S|(u) * |S|(v)) = E(|S|(u * v)) = E(u * v)$ , giving (i). Using this, for any words  $u, v \in \mathbb{A}^*$  we have  $E((\pi_{g=n} \circ S)(u) * (\pi_{g=n} \circ S)(v)) = (-1)^{2n} E(|S|(\pi_{g=n} \circ u) * |S|(\pi_{g=n} \circ v)) = E((\pi_{g=n} \circ u) * (\pi_{g=n} \circ v))$ , giving (ii).  $\square$

Finally, our main result is as follows.

**Theorem 4 (Main result: Efficiency of the antisymmetric sign reverse integrator)** *Consider the flowmap of a stochastic differential equation driven by independent Lévy processes with moments of all orders. Assume the flowmap possesses a separated stochastic Taylor expansion, and that the vector fields are sufficiently smooth to ensure convergence of the truncated stochastic Taylor integration schemes to all orders. The antisymmetric sign reverse approximation at a given truncation level  $n$  is efficient in the sense that its local leading order mean-square errors are always smaller than those of the truncated stochastic Taylor scheme of the same order, independent of the driving vector fields and of the initial conditions. It converges with global strong mean-square order of convergence  $n$ .*

*Proof* It suffices to show the leading order remainder endomorphism  $\pi_{g=n+1} \circ \text{id}$  associated with the truncated stochastic Taylor expansion has greater norm than the remainder endomorphism  $\pi_{g=n+1} \circ \frac{1}{2}(\text{id} - S)$  associated with the antisymmetric sign reverse integrator, independent of  $\mathbf{V}$ . Using Lemma 4 and setting  $\pi = \pi_{g=n+1}$ , we have

$$\begin{aligned} & \|\pi \circ \text{id}\|^2 \\ &= \left\| \pi \circ \left( \frac{1}{2}(\text{id} - S) + \frac{1}{2}(\text{id} + S) \right) \right\|^2 \\ &= \left\| \pi \circ \frac{1}{2}(\text{id} - S) \right\|^2 + \left\langle \pi \circ \frac{1}{2}(\text{id} - S), \pi \circ \frac{1}{2}(\text{id} + S) \right\rangle + \left\| \pi \circ \frac{1}{2}(\text{id} + S) \right\|^2 \\ &= \left\| \pi \circ \frac{1}{2}(\text{id} - S) \right\|^2 + \frac{1}{4} \left( \|\pi \circ \text{id}\|^2 - \|\pi \circ S\|^2 \right) + \left\| \pi \circ \frac{1}{2}(\text{id} + S) \right\|^2 \\ &= \left\| \pi \circ \frac{1}{2}(\text{id} - S) \right\|^2 + \left\| \pi \circ \frac{1}{2}(\text{id} + S) \right\|^2, \end{aligned}$$



We observe  $\|\pi_{g=n+1} \circ \frac{1}{2}(\text{id} - S)\|^2 \leq \|\pi_{g=n+1} \circ \text{id}\|^2$ , giving the result. Strong global mean-square order of convergence  $n$  for the scheme is a natural consequence.  $\square$

## 6 Numerical simulations

We present two numerical experiments demonstrating the results proved above. First, we study the linear constant coefficient drift-diffusion equation  $dy_t = A_0 y_t dt + A_1 y_t dW_t^1 + A_2 y_t dW_t^2$ , where  $y_t \in \mathbb{R}^2$ , the  $W^i$  are independent Wiener processes, the initial condition is  $y_0 = (1, 0.5)^T$  and the coefficients are given by the matrices

$$A_0 = \begin{pmatrix} -0.0721 & -0.2173 \\ -0.1719 & -0.9581 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0.0800 & 0.5769 \\ -0.5961 & -0.9619 \end{pmatrix}$$

and  $A_2 = \begin{pmatrix} -0.9438 & 0.5520 \\ -0.4684 & 0.1591 \end{pmatrix}.$

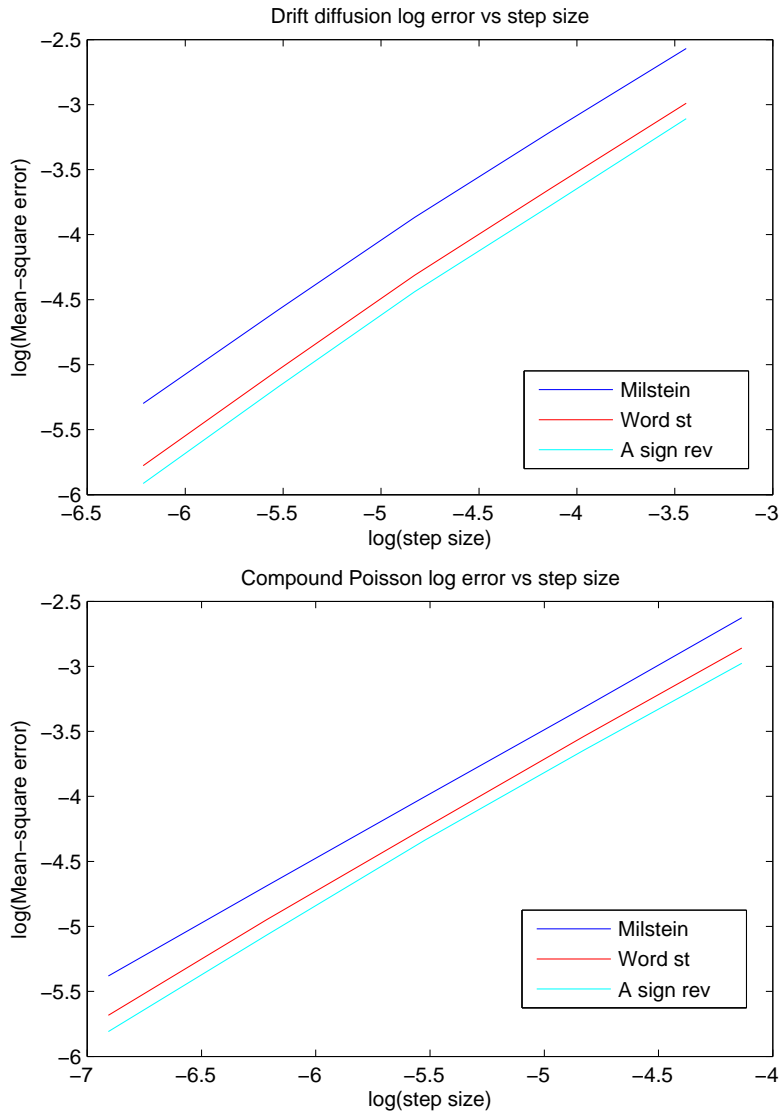
We compare the global mean-square error  $E(\sup_{0 \leq t \leq T} |y_t - \hat{y}_t|^2)^{1/2}$ , estimated by sampling 10000 paths, for three different numerical approximations  $\hat{y}_t$ : (i) the Milstein scheme, (ii) the order 1 word length truncated stochastic Taylor expansion, and (iii) the order 1 word length antisymmetric sign reverse scheme. Figure 1 (top panel) shows the global mean-square error of the numerical approximations against the discretization grid stepsize. In line with our analytical results, the antisymmetric sign reverse integrator gives an improvement over the word length stochastic Taylor scheme, which is in turn more accurate than the Milstein scheme.

We now consider the linear constant coefficient equation  $dy_t = A_0 y_t dt + A_1 y_t dW_t + A_2 y_t dJ_t$ , where  $W_t$  is a Wiener process and  $J_t$  is the compound Poisson process  $J_t = \sum_{i=1}^{N_t} v_i$ . Here  $N_t$  is a standard Poisson process with intensity  $\lambda$  and independent, identically distributed marks  $v_i \sim \mathcal{N}(0, \lambda^{-1})$ . The initial condition is again  $y_0 = (1, 0.5)^T$ . We compare the same three schemes: (i) the Milstein scheme, (ii) the order 1 word length truncated stochastic Taylor expansion, and (iii) the order 1 word length antisymmetric sign reverse scheme. For our simulations, we have taken  $\lambda = 43.40$ , and matrices

$$A_0 = \begin{pmatrix} -0.6841 & 0.5588 \\ 0.9335 & -0.0273 \end{pmatrix}, \quad A_1 = \begin{pmatrix} 0.9530 & -0.6357 \\ 0.9556 & -0.3250 \end{pmatrix}$$

and  $A_2 = \begin{pmatrix} -0.1763 & -0.9088 \\ 0.3382 & 0.8925 \end{pmatrix}.$

We observe in Figure 1 (bottom panel) that the antisymmetric sign reverse integrator is more accurate than the word length stochastic Taylor integrator, which in turn is more accurate than the Milstein scheme.



**Fig. 1** For a linear drift-diffusion (top panel) and linear jump-diffusion (bottom panel) stochastic differential equations, we see that, with word length grading, the order 1 anti-symmetric sign reverse integrator is more accurate than the corresponding stochastic Taylor scheme, which in turn is more accurate than the Milstein scheme (which is determined by mean-square grading).

## 7 Global convergence

The error estimates derived in §4 are local estimates. We require a means of obtaining global convergence from local error estimates. In the case of drift-diffusion equations, such a mechanism was established by Milstein [26, 27]. In this section we present a natural generalization of Milstein's theorem to equations driven by Lévy processes. The proof of Milstein's theorem requires the following bounds and continuity results for the exact flow. The proofs are standard and are not reproduced here, for details see Situ [37, p. 76–78], Fujiwara & Kunita [16, p. 84–86], Applebaum [2, p. 332–336].

**Lemma 5 (Continuity and bounds for flows of Lévy-driven equations)** *Let  $\varphi_{s,t}$  be the flow of a Lévy-driven equation. For any  $\mathcal{F}_s$ -measurable  $x, y \in L^2$ , and  $s < t$  in  $[0, T]$ , there exists a constant  $K > 0$  such that: (i)  $\|\varphi_t(y)\|_{L^2}^2 \leq K(1 + \|y\|_{L^2}^2)$ ; (ii)  $\|\varphi_{s,t}(y) - y\|_{L^2}^2 \leq K(t - s)(1 + \|y\|_{L^2}^2)$ ; (iii)  $\|\varphi_t(x) - \varphi_t(y)\|_{L^2}^2 \leq e^{Kt}\|x - y\|_{L^2}^2$  and (iv)  $\|\varphi_{s,t}(x) - \varphi_{s,t}(y) - (x - y)\|_{L^2}^2 \leq K(t - s)\|x - y\|_{L^2}^2$ .*

We now present the generalization of Milstein's theorem to Lévy-driven equations. Essentially, it states that any Markovian integration scheme for which the local mean-square error converges with order  $p$ , converges globally in the mean-square sense with order  $p - 1/2$ , provided the expectation of the local error is at least of order  $p + 1/2$ .

**Theorem 5 (Generalized Milstein Theorem)** *Let  $\hat{\varphi}_{s,t}$  be an approximate flow defined for values  $s \leq t$  such that  $s, t \in \{h, 2h, \dots, T\}$ , where  $h$  is the step size of the scheme. Suppose that the expectation of the local error of the approximation is of order  $p_1$ , and that the local mean-square error is of order  $p_2$ , i.e. for any  $t < T$  there exists a constant  $K$  such that*

$$\begin{aligned} \left| E(\varphi_{t,t+h}(y) - \hat{\varphi}_{t,t+h}(y)) \right| &\leq K(1 + |y|^2)^{1/2} h^{p_1}, \\ \|\varphi_{t,t+h}(y) - \hat{\varphi}_{t,t+h}(y)\|_{L^2} &\leq K(1 + |y|^2)^{1/2} h^{p_2}, \end{aligned}$$

where  $p_2 \geq 1/2$  and  $p_1 \geq p_2 + 1/2$ . Suppose further that  $\hat{\varphi}_{s,t}$  is independent of  $\mathcal{F}_s$  for all  $s < t$ . Then the approximation converges globally to the exact flow with strong order  $p_2 - 1/2$ , i.e. there exists a constant  $K$  such that for all  $t \in \{h, 2h, \dots, T\}$  we have  $\|\varphi_t(y_0) - \hat{\varphi}_t(y_0)\|_{L^2} \leq K(1 + \|y_0\|_{L^2}^2)^{1/2} h^{p_2 - 1/2}$ .

The proof given in Milstein [26, 27] for equations driven by Wiener processes relies only on the local accuracy of the approximation and the continuity and growth bounds for the exact flow given in the preceding lemma. By Lemma 5, these properties of the exact flow hold for Lévy-driven equations, and hence the proof of Milstein immediately generalizes to give the above theorem. We remark that the assumption that  $\hat{\varphi}_{s,t}$  is independent of  $\mathcal{F}_s$  is required, as the local error estimates must hold across each computational interval  $[s, t]$  where the expectation is taken with respect to  $\mathcal{F}_s$ . All the schemes we consider fulfil this property.

## Acknowledgement

C.C. would like to thank Michael Tretyakov and Seva Shneer for useful comments.

## References

1. Abe, E. 1980 *Hopf algebras*, Cambridge University Press.
2. Applebaum, D. 2004 *Lévy Processes and Stochastic Calculus*, Cambridge University Press.
3. Barndorff-Nielsen, O., Mikosch, T., Resnick, S., ed. 2001 *Lévy processes: theory and applications*, Springer, New York.
4. Castell, F., Gaines, J., 1995 An efficient approximation method for stochastic differential equations by means of the exponential Lie series, *Math. Comp. Simulation* **38**, 13–19. (doi:10.1016/0378-4754(93)E0062-A)
5. Castell, F., Gaines, J., 1996 The ordinary differential equation approach to asymptotically efficient schemes for solutions of stochastic differential equations, *Ann. Inst. H. Poincaré Probab. Statist.* **32**, 231–250.
6. Chen, K.-T. 1968 Algebraic paths., *Journal of Algebra*, **10**, 8–36.
7. Clark, J.M.C. 1982 An efficient approximation for a class of stochastic differential equations, in *Advances in Filtering and Optimal Stochastic Control, Cocoyoc, Mexico, 1982, Lecture Notes in Control and Information Sciences*, **42**, Berlin etc: Springer. (doi:10.1007/BFb0004526)
8. Cont, R., Tankov, P. 2004 *Financial Modelling with Jump Processes*, Chapman & Hall/CRC.
9. Curry, C., 2014 *Algebraic structures in stochastic differential equations*, PhD thesis, Heriot-Watt University.
10. Curry, C., Ebrahimi-Fard, K., Malham, S.J.A., Wiese, A. 2014 Lévy processes and quasi-shuffle algebras, *Stochastics* **86**(4), 632–642. (doi:10.1080/17442508.2013.865131)
11. Ebrahimi-Fard, K., Lundervold, A., Malham, S.J.A., Munthe-Kaas, H., Wiese, A. 2012 Algebraic structure of stochastic expansions and efficient simulation, *Proc. R. Soc. A* **468**, 2361–2382. (doi:10.1098/rspa.2012.0024)
12. Ebrahimi-Fard, L., Malham, S.J.A., Patras, F., Wiese, A., 2015 Flows and stochastic Taylor series in Itô calculus, *J. Phys. A*, **48**, 495202 (17pp). (doi:10.1088/1751-8113/48/49/495202)
13. Ebrahimi-Fard, L., Malham, S.J.A., Patras, F., Wiese, A., 2015 The exponential Lie series for continuous semimartingales, *Proc. R. Soc. A* **471**. (doi:10.1098/rspa.2015.0429)
14. Eilenberg, S, Mac Lane, S. 1953. On the groups  $H(\pi, n)$ . *Annals of Mathematics* **58**(1), 55–106.
15. Friz, P., Shekhar, A. 2012 General Rough Integration, Lévy Rough Paths and a Lévy-Kintchine type formula, *arXiv:1212.5888*.
16. Fujiwara, T., Kunita, H. 1985 Stochastic differential equations of jump type and Lévy processes in diffeomorphisms group, *J. Math. Kyoto Univ.* **25**(1), 71–106.
17. Hoffman, M.E. 2000 Quasi-shuffle products, *Journal of Algebraic Combinatorics* **11**, 49–68. (doi:10.1023/A:1008791603281)
18. Hoffman, M.E., Ihara, K. 2012 Quasi-shuffle products revisited, Preprint. Max-Planck-Institut für Mathematik Bonn.
19. Gaines, J. 1994 The algebra of iterated stochastic integrals, *Stochast. and Stochast. Rep.* **49**, 169–179. (doi:10.1080/17442509408833918)
20. Kloeden, P.E., Platen, E. 1999 *Numerical Solution of Stochastic Differential Equations* (3rd Printing), Berlin etc: Springer.
21. Kloeden, P.E., Platen, E., Wright, I. 1992 The approximation of multiple stochastic integrals, *Stochastic Anal. Appl.* **10**(4), 431–441. (doi:10.1080/07362999208809281)
22. Li, C.W., Liu, X.Q. 1997 Algebraic structure of multiple stochastic integrals with respect to Brownian motions and Poisson processes, *Stochastics and Stoch. Reports* **61**, 107–120.

23. Lord, G., Malham, S.J.A. & Wiese, A. 2008 Efficient strong integrators for linear stochastic systems, *SIAM J. Numer. Anal.* **46**(6), 2892–2919.
24. Malham, S.J.A., Wiese, A. 2008 Stochastic Lie group integrators, *SIAM J. Sci. Comput.* **30**, 597–617. (doi:10.1137/060666743)
25. Malham, S.J.A., Wiese, A. 2009 Stochastic expansions and Hopf algebras, *Proc. R. Soc. A* **465**, 3729–3749. (doi:10.1098/rspa.2009.0203)
26. Milstein, G.N. 1987 A theorem on the order of convergence of mean-square approximations of solutions of systems of stochastic differential equations, *Theor. Prob. Appl.* **32**(4), 738–741.
27. Milstein, G.N. 1995 *Numerical integration of stochastic differential equations*, Mathematics and Its Applications **313**, Kluwer Academic Publishers.
28. Newton, N.J. 1986 An asymptotically efficient difference formula for solving stochastic differential equations, *Stochastics*, **19**, 175–206 (doi:10.1080/17442508608833423)
29. Newton, N.J. 1991 Asymptotically efficient Runge-Kutta methods for a class of Itô and Stratonovich equations, *SIAM J. Appl. Math.* **51**, 542–567. (doi:10.1137/0151028)
30. Platen, E. 1980 Approximation of Itô integral equations, In *Stochastic differential systems*, Volume 25 of *Lecture notes in Control and Inform. Sci.*, 172–176. Springer. (doi:10.1007/BFb0004008)
31. Platen, E. 1982 A generalized Taylor formula for solutions of stochastic differential equations, *SANKHYA A* **44**(2), 163–172.
32. Platen, E., Bruti-Liberati, N. 2010 *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Berlin etc: Springer.
33. Platen, E., Wagner, W. 1982 On a Taylor formula for a class of Itô processes, *Probab. Math. Statist.* **3**(1), 37–51.
34. Protter, P. 2003 *Stochastic Integration and Differential Equations (2nd ed.)* Berlin, etc: Springer.
35. Radford, D.E. 1979 A natural ring basis for the shuffle algebra and an application to group schemes, *Journal of Algebra* **58**, 432–454. (doi:10.1016/0021-8693(79)90171-6)
36. Reutenauer, C. 1993 *Free Lie Algebras*, London Mathematical Society Monographs, Clarendon Press, Oxford.
37. Situ, R. 2005 *Theory of Stochastic Differential Equations with Jumps and Applications: Mathematical and Analytical Techniques with Applications to Engineering*, Springer.
38. Schützenberger, M.-P. 1958. Sur une propriété combinatoire des algèbres de Lie libres pouvant être utilisée dans un problème de mathématiques appliquées., *Séminaire P. Dubreil. Algèbre et théorie des nombres* **12**(1), 1–23.
39. Strichartz, R.S. 1987 The Campbell–Baker–Hausdorff–Dynkin formula and solutions of differential equations, *J. Funct. Anal.* **72**, 320–345. (doi:10.1016/0022-1236(87)90091-7)
40. Sussmann, H.J. 1986 A product expansion for the Chen series, *Theory and Applications of Nonlinear Control Systems*, C.I. Byrnes and A. Lindquist (editors), Elsevier Science Publishers B.V. (North Holland).